**Project Proposal**

|  |  |
| --- | --- |
| **Name:** | Legie Grieve |
| **Programme:** | MSc Data Science |
| **Module ID:** | UFCF9Y-60-M CSCT Masters Project |
| **Student ID:** | 21051647 |
| **Project Title:** | Investigating Patterns of Depression Symptoms Among Young Adults Impacted by the COVID-19 Pandemic Using Association Rule Learning |

**Problem Statement:**

The COVID-19 pandemic, declared by the World Health Organization in March 2020, led to widespread social isolation due to lockdowns, social distancing measures, and disruptions to daily life. This unprecedented level of isolation significantly impacted mental health across the globe, with a notable increase in depression symptoms among various populations. However, the patterns and interrelationships among these depression symptoms remain underexplored to some extent.

**Aim and scope:**

This study aims to investigate the patterns of depression symptoms among young people aged 18 to 25 years old by applying Association Rule Learning techniques. By identifying frequently co-occurring symptoms and uncovering underlying associations, this research seeks to provide deeper insights into how depression manifests during the pandemic.

**Objectives:**

1. To perform Exploratory Data Analysis (EDA) to uncover trends and patterns within the dataset.
2. To identify correlated depression symptoms among young people age from 18 to 25 years during global COVID-19 pandemic using Association Rule Learning.

**Expected Outcomes:**

**Insightful Trends and Patterns:**

Through Exploratory Data Analysis (EDA), the study is expected to uncover significant trends and patterns within the dataset, providing a clearer understanding of how depression symptoms have manifested during the COVID-19 pandemic.

**Identification of Key Symptom Associations:**

By applying Association Rule Learning to the data, the study aims to identify specific combinations of depression symptoms that are strongly correlated among young people aged 18 to 25 during the global COVID-19 pandemic.

**Scope of the Study:**

**Population Focus:**

The study will focus specifically on young adults aged 18 to 25 years who experienced the COVID-19 pandemic. This demographic is chosen due to prevalence of mental health issues during the pandemic, making it a critical group for understanding the impact of social isolation on mental health.

**Variables:**

**Age:**

This variable specifically targeting young adults aged 18 to 25 years. Age is a crucial factor in understanding how depression symptoms and their patterns may differ among different age group.

**Gender:**

Gender is included to explore any differences in depression symptoms between male, female and other individuals within the young adult age group.

**Depression Symptoms:**

Depression symptoms will be selected based on an extensive review of previous research studies on the issue.

**Data Source:**

For this study, synthetic data will be created to simulate the depression symptoms experienced by young adults during the COVID-19 pandemic. This synthetic dataset will be designed to mirror the characteristics and patterns observed in real-world data, allowing for a comprehensive analysis despite the absence of actual data.

**Method:**

To explore and identify patterns in depression symptoms, Association Rule Learning (ARL) was used. ARL is an unsupervised machine learning technique designed to uncover detailed relationships between variables, showing how different depression symptoms are related to one another.

**Ethical considerations:**

Since the data is generated synthetically, there is no ethical issue to adhere to.

**Out of Scope:**

* The study does not include a longitudinal analysis of symptom patterns over time.
* The research will not compare depression symptoms with other mental health conditions or explore symptoms beyond those related to depression.
* This study will not involve the deployment or implementation of the model to the real-world data setting. The focus is strictly on analysis and identifying patterns within the data.

**Brief review of relevant literature:**

When WHO declared Covid-19 a pandemic in 2019, numerous nations implemented a range of restrictions to curb the spread of virus. These measures include social distancing, wearing mask, stay-at-home orders, remote work arrangements, and closure of non-essential services, with exceptional societal and economic repercussions. (Pennix et al, 2022)

“The Coronavirus Disease 2019 (COVID-19) pandemic has threatened global mental health, both indirectly via disruptive societal changes and directly via neuropsychiatric sequelae after SARS-CoV-2 infection.” (Pennix et al, 2022)

*“Psychological distress due to the pandemic was increased by the reduction in social contacts, forced isolation, and quarantine” (de girolamo et al. 2022)*

During the first year of the COVID-19 pandemic, the global rates of anxiety and depression increased by 25%, as reported in a scientific brief by the WHO.

Nochaiwong, S., Ruengorn, C., Thavorn, K. *et al.* reported that, during the pandemic, the global prevalence of mental health issues among the general population increased significantly. Their research included data from 32 different countries and 398,771 participants, highlighting varied prevalence rates across regions. They found the prevalence estimates to be 28% for depression, 26.9% for anxiety, 24.1% for post-traumatic stress symptoms, 36.5% for stress, 50% for psychological distress, and 27.% for sleep problems.

There have been numerous analyses conducted to visualise prevalent mental health issues, yet relatively little research has focused on uncovering associations between various mental health issues. Association Rule Learning is an unsupervised machine learning technique that will help to identify the corelated issues within dataset. It is widely used data mining technique that identify interesting patterns within datasets. Originally developed for market basket analysis, this method has proven versatile and can be effectively applied across various domains. (Kost et al, 2012)

There are two main steps in the process of ARL:

1. Frequent itemset mining: Identify groups of items that commonly appear together in the dataset, according to a predefined minimum support threshold.
2. Rule generation: Create association rules from the frequent itemset, using predefined minimum confidence threshold. (Anon, 2023).

The common metrics used in ARL are Support, Confidence, and Lift and most common types of ARL are Apriori, ECLAT and FP Growth.

**Project plan:**

The project is scheduled to be completed within four months, starting from May 1st and ending August 29th, 2024. To manage the project efficiently, a project design consisting of seven steps has been created, as detailed in Appendix 1. The timescale breakdown is provided in Appendix 2.

**Appendices**

**Appendix 1**

**Project Design**

A screenshot of a diagram

Description automatically generated

**Appendix 2**

**Time Scale (Gantt Chart)**

A screenshot of a computer

Description automatically generated

**References:**

1. Penninx, B.W.J.H., Benros, M.E., Klein, R.S. *et al.* (2022). How COVID-19 shaped mental health: from infection to pandemic effects. *Nat Med* **28**, 2027–2037. [*https://doi.org/10.1038/s41591-022-02028-2*](https://doi.org/10.1038/s41591-022-02028-2) [online]. [Accessed 18 March 2024].
2. World Health Organisation, (2022) Mental Health and Covid-19: Early Evidence of the Pandemic’s Impact. Https://iris.who.int/bitstream/handle/10665/352189/who-2019-ncov-sci-brief-mental-health-2022.1-eng.pdf?sequence=1 [online]. [Accessed 18 March 2024].
3. Rume, T. and Islam, S.M. (2020) Environmental Effects of Covid-19 Pandemic and Potential Strategies of Sustainability. Https://www.ncbi.nlm.nih.gov/pmc/articles/pmc7498239/ [online]. [Accessed 18 March 2024].
4. Eslami, H. and Jalili, M. (2020) The Role of Environmental Factors to Transmission of Sars-cov-2 (Covid-19). Https://www.ncbi.nlm.nih.gov/pmc/articles/pmc7226715/ [online]. [Accessed 18 March 2024].
5. World Health Organisation, (2022) Covid-19 Pandemic Triggers 25% Increase in Prevalence of Anxiety and Depression Worldwide. Https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide [online]. [Accessed 18 March 2024].
6. British Broadcast Corporation, (2021) How Has Coronavirus Affected Mental Health?. https://www.bbc.co.uk/news/explainers-55957105 [online]. [Accessed 21 November 2023].
7. Office For National Statistics, (2020) Coronavirus and Depression in Adults, Great Britain. Https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/coronavirusanddepressioninadultsgreatbritain/june2020 [online]. [Accessed 21 November 2023].
8. Kupcova, I., Danisovic, L., Klein, M. and Harsanyi, S. (2023) Effects of the Covid-19 Pandemic on Mental Health, Anxiety, and Depression. Https://bmcpsychology.biomedcentral.com/articles/10.1186/s40359-023-01130-5 [online]. [Accessed 21 November 2023].
9. Kost R, Littenberg B, Chen E, (2012), Exploring Generalized Association Rule Mining for Disease Co-Occurrences, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540474/> [Online] [Accessed 11 May 2024]
10. Noble J, (2024) What is the Apriori algorithm? <https://www.ibm.com/topics/apriori-algorithm> [Online] [Accessed 05 July 2024]
11. Anon (2024), ML ECLAT, <https://www.geeksforgeeks.org/ml-eclat-algorithm/> [Online] [Accessed 05 July 2024]
12. Anon (2024) Frequent Pattern Growth Algorithm, <https://www.geeksforgeeks.org/frequent-pattern-growth-algorithm/> [Online] [Accessed 05 July 2024]